

Exploring the extent of openness of open government data – A critique of open government datasets in the UK

Victoria Wang and David Shepherd

Abstract

The claim of the open government data (OGD) movement is that it would add considerable value to the political, economic and social development of nations. Thus, during the past few years, governments around the world have been under increasing pressure to release data to encourage citizen participation in government. In the UK, the government publishes its data on the data.gov.uk website for the public to find, view, download and interrogate.¹ In this paper, we evaluate the extent of openness of OGD in the UK – a recognised global leader in the OGD movement – by examining a sample of 400 datasets listed on this website. The examination uses an ‘ordinary citizen’ test, which is developed based on the original eight Sebastopol principles of open data. Our examination discovers that existing prevalent measures of openness of data are inadequate. Our findings demonstrate that the majority of the published government resources on the website are informational rather than granular data. In fact, only a small minority of these advertised as open data are actually open, which severely limits their open government utilities.

Keywords: Open government data, open data, open government, data.gov.uk, ordinary citizen, granularity

¹ See: <https://data.gov.uk/>.

1. Introduction

Open government data (OGD) is a sub-set of open data and refers to government-related data that is opened to the public in order to support and enable the grand, democratic purposes of open government (Kučera et al., 2013). The movement seeks radical democratic transformation by leveraging the value of data for society through three principles: i) openness, ii) participation, and iii) collaboration (Davies, Perini and Alonso, 2013; White House, 2009). Thus, the opening of government data is not the ultimate objective, rather it is a very important democratic mechanism to promote citizen engagement in challenging and assisting open governments (Davies and Perini, 2016; McGee and Edwards, 2016; Schrock and Shaffer, 2017). OGD emerged as an important, multilateral movement among state administrations after the turn of the 21st century. It represents a paradigm shift in the way that governments shape their relationships with the public (Gigler, Custer and Rahmetulla, 2011). A seminal event in its development was the publication of eight principles of OGD following a meeting of thirty open government activists in Sebastopol in 2007. These eight principles, known as the Sebastopol principles, are i) complete, ii) primary, iii) timely, iv) accessible, v) machine-processable, vi) non-discriminatory, vii) non-proprietary, and viii) licence free.² A second key milestone was the signing of the G8 Open Data Charter by the leaders of the G8 group in 2013 (G8, 2013). The aims of this Charter are clearly stated in its five principals: i) open data by default; ii) quality and quantity; iii) useable by all; iv) releasing data for improved governance; and v) releasing data for innovation (G8, 2013). The Charter requires the signatory governments to publish freely accessible, granular open data which is machine-readable for i) data mining, ii) linking, and iii) in-depth analysis.

² More information on the eight Sebastopol principles can be found at <https://opengovdata.org/>.

The UK government is one of the original signatories, and has demonstrated its continued commitment to the G8 Open Data Charter by a series of activities, such as its publishing of the ‘Open Government National Action Plan 2016 to 2018: mid-term self assessment’ (HM Government, 2016); its membership of the Open Government Partnership;³ iii) its top position on the Open Data Barometer;⁴ and especially, its release of government data on the data.gov.uk website to ordinary citizens.⁵ This website is a portal that has catalogued data published by central government, local authorities, government owned corporations and regulators in the UK since 2010.⁶ As indicated by its slogan – ‘Opening up Government’ – the site’s home page describes its purpose as, ‘Find data published by government departments and agencies, public bodies and local authorities. You can use this data to learn more about how government works, carry out research or build applications and services’.⁷ It makes a symbolic claim about providing access to raw data that drives government.⁸ Of course, it is unlikely that the online data repository contains all the open data published by government bodies as it relies on the diligence of local data custodians to publish data files on this website. Nevertheless, it contains the largest, most representative collection of open data available in the UK. As of the 1st November, 2017, data.gov.uk contained 42,991 datasets. Nevertheless, this does not mean that the OGD movement in the UK is actually fulfilling its grand purpose of advancing open government. Conceptual ambiguities associated with the OGD movement provide the space for governments to manoeuvre, even to the extent of exaggerating their open government credentials

³ See: <https://www.opengovpartnership.org/participants>.

⁴ See: <https://opendatabarometer.org/>.

⁵ See: <https://data.gov.uk/>.

⁶ Note that it also includes references to unpublished data, although it does not usually justify why it is not published or provide any advice on how to access to it.

⁷ See: <https://data.gov.uk/>.

⁸ See: <https://data.gov.uk/about/>.

– OGD can be exploited to enable the impression of open government (Heimstädt, 2017; Kornberger, 2017).

In this paper, we evaluate the extent of openness of open government datasets in the UK. In general, there are three important dimensions in assessing the contribution of an OGD programme to open government. These are: i) the extent of the openness of published materials; ii) the value to open government of the subject matter contained in these materials; and iii) whether the government actually engages with feedback. The extent of openness is a more objective dimension that is concerned with the number and proportion of materials released under the open data label that actually comply with the open data definition. The subject value dimension is concerned with the value of topics in enabling accountability and novel insights into the policies and performances of governments. Some data, such as the locations of public toilets or traffic information, have little utility in furthering open government but allow authorities to promulgate the illusion of open government (Grimmelikhuijsen, 2013; Janssen, 2012; Kornberger et al., 2017). The feedback engagement dimension is crucial since transparency is not the same as listening or accountability (Davies and Perini, 2016; Janssen et al., 2012). Our evaluation focuses on the first dimension, and asks the overarching research question – what is the extent of openness of OGD released by the UK government?

We examine a sample of 400 datasets published on the centralised government portal – data.gov.uk – to estimate the proportion of materials that are actually open data. The portal uses the Comprehensive Knowledge Archive Network (CKAN) platform, which is widely deployed globally to manage open data websites.⁹ This research does not evaluate the performance of the data.gov.uk portal; it uses the portal as a convenient means to access UK OGD datasets in order

⁹ See: <https://ckan.org>.

to evaluate their openness. We conduct a review of definitions of OGD and its related assessment frameworks, and discover that existing prevalent measures of openness of data are inadequate for our evaluation. Thus, we use an ‘ordinary citizen’ test, which is developed based on the original eight Sebastopol principles, and inspired by Lourenço (2015)’s proposal of viewing data from the perspective of the ‘ordinary citizen’. Two samples were selected from the data.gov.uk: i) a purposive sample of the 136 most popular datasets; and ii) a random sample of 264 datasets. Our findings suggest that just one in ten open government data resources actually contain open data. The lack of granularity (aggregation) is the most common reason that the overwhelming majority of datasets fail the ordinary citizen test – 65% of datasets contained aggregate ‘open information’ rather than granular ‘open data’.

The UK’s position as the leader of the OGD movement globally and our discovery of the general weaknesses in prevalent measures of openness, mean that our findings have relevance to the broader OGD movement. This paper proceeds with Section 2 presenting a review of current literature on various aspects of OGD, including its purposes, delivery, technical attributes, framework definitions, and current openness assessment frameworks. Section 3 sets out the research method. This is followed by the main findings in Section 4. Section 5 presents the discussions of findings, which includes a critique of the prevalent openness measures. Finally, Section 6 concludes the paper.

2. What is open government data?

Currently, there is no single, universally accepted definition of *open data* (Gigler, Custer and Rahemtulla, 2011; Kitchin, 2014; Verhulst and Young, 2016). As a result, *open data* and *open government data* are often conceptually bound with the grand purposes of *open government*

(McGee, 2011; McGee and Gaventa, 2011; Yiu, 2012). In fact, this has created a conceptual ambiguity among open data, open government data, and open government. This is further complicated by the loose interpretation of the term ‘*data*’ to encompass anything from raw, granular datasets to written reports. The conceptual ambiguity hinders progress in developing practical and theoretical understandings of the OGD movement (Davies and Perini, 2016; McGee and Edwards, 2016). Crucially, this ambiguity allows closed governments to claim open governance credentials simply by building “the right kind of website” (Yu and Robinson, 2012, p. 179) or by publishing a sufficient weight of government material, irrespective of their value to transparency, accountability and open governance (Grimmelikhuijsen, 2013; Janssen et al. 2012; Kornberger et al., 2017).

2.1 Purposes of OGD

Some definitions of OGD focus on its purpose and its key role in enabling opening government (Kučera et al., 2013). The preamble to the Open Data Charter clearly expresses the grand purposes of OGD in seeking technology fuelled innovation that creates more accountable, efficient, responsive, and effective governments and businesses, and spurs economic growth (G8, 2013). The UK government, for example, describes OGD as “...the building block for a more democratic, equal and sustainable society (HM Government, 2016, p. 3). This bandwagon is driven by the spirit of democracy and economy (Chan, 2013) that empowers ordinary citizens (Baack, 2015; Parycek and Sachs, 2010), guards against the most egregious forms of political corruption (Attard et al., 2015; Ballingall, 2011), and stimulates innovations for government efficiencies (Millard, 2015) and private sector growth (Janssen et al., 2012). In 2015, a European Commission report, for example, forecasted that OGD would create €76 billion for the European annual market, €1.7 billion in government efficiency savings, and 350,000 new

jobs by 2020, whilst acknowledging that there is limited evidence of any actual benefit to date (European Commission, 2015).

2.2 Enabling delivery of the purposes

Other definitions of OGD are concerned with the means of delivering its grand purpose of open government, emphasising accessibility and the freedom for citizens to use OGD. The Open Data Institute (ODI) states that: “Open data is data that anyone can access, use or share. Simple as that.”¹⁰ Open Knowledge International states that: “Open data and content can be freely used, modified, and shared by anyone for any purpose”.¹¹ Opening up government data is crucial in freely distributing the ability to produce socially useful statistics, information and knowledge (Kitchin, 2015; Open Knowledge International;¹² Parycek and Sachs, 2010). Tauberer (2014) characterises this form of citizen-participation in open government as ‘civic hacking’, which he understands as the use of OGD to make governments more accountable, and to find creative solutions to interesting problems. Eaves (2009) succinctly characterises OGD as that which enables users to ‘find, play, share’. Discoverability and accessibility are thus prerequisites and without these two attributes, government data cannot be regarded as open (Attard et al., 2015). One approach is to attach metadata to datasets (Berners-Lee, 2009; Kubler et al., 2017). Another is the development of centralised government portals, like the UK’s data.gov.uk, which can be used by “...even ordinary citizens, without specialized technical skills...” (Lourenço, 2015, p. 324). To realise such sharing, the data must not be burdened with legal barriers, such as copyright, patent and licence restrictions (Chatfield and Reddick, 2017). Open Knowledge International regards licence conditions as a central requirement of open data. In practice these

¹⁰ See: <https://theodi.org/what-is-open-data>.

¹¹ See: <http://opendefinition.org>.

¹² See: <https://okfn.org/about/>.

legal impediments are generally addressed by releasing data that are either licence-free (G8, 2013) or freely usable under an open licence (Thorsby et al., 2017).

2.3 Technical attributes of OGD

Essential attributes of OGD that directly support the delivery of open government and economy are that it is unmediated, granular, structured, machine-readable data (Robinson et al., 2009). Other attributes extend the accessibility and usability of the data: an open, non-proprietary format ensures that data is accessible to anyone with a computer with the need to purchase proprietary software; metadata and data interchange features improve the discoverability and usability of the data by attaching machine-readable descriptions of the data using programming languages such as the Resource Description Framework (RDF) (Attard et al., 2015; W3C, 2004). However, these additional attributes are irrelevant if the data is not granular and machine-readable. Yu and Robinson (2012) situated data on an adaptability spectrum: machine-readable, granular data has more value due to its adaptability, whilst aggregated PDF information is less valuable as it is inert. The data, information, knowledge, wisdom hierarchy (DIKW) clearly distinguishes between data and information (Rowley, 2007).

Most of the existing definitions of open data accommodate digital data released by anyone, about any topic, including government, social and scientific data (Kitchin, 2013; Lagoze, 2014; Leonelli, 2012; Mayernik, 2017). Descriptions that focus on the technical aspects of datasets are often silent about the usefulness of the data. It is, however, recognised that legal, moral, security and privacy restrictions are barriers to useful data (Janssen and van den Hoven, 2015). Janssen et al. (2012, p. 259) elegantly define OGD as “... non-privacy-restricted and non-confidential data which is produced with public money and is made available without any restrictions on its usage or distribution”. The privacy barrier is a key OGD challenge (Axelsson

and Schroeder, 2009) that can limit access to important data (Ruijter et al., 2017). In the UK, the General Data Protection Regulations (GDPR) and the Statistics and Registration Act 2009 limit the publication and use of personal data. In practice, this often means that data is aggregated into categories or summaries in order to hide the detail (Altman et al., 2015). Under these circumstances, the data is no longer open data, it has become mediated open information (cf. Rowley, 2007; Tuomi 1999).

2.4 Framework definitions

A number of framework definitions have been created, which variously list specific attributes of government datasets and organisational requirements. An early, foundational development was the eight Sebastopol principle in 2007.¹³ These principles provide a roadmap for publishers and are the basis of most guidelines for publishing OGD on the web (Attard et al., 2015):

1. *Complete* - all public data, which is not subject to privacy, security or privilege limitations, is made available.
2. *Primary* - data is granular and made available as it is available at the source, not aggregated or modified.
3. *Timely* - data is made available to the public as soon as possible after the actual data is created, in order to preserve the value of the data.
4. *Accessible* - data is made available to all consumers possible, and with no limitations on its use.
5. *Machine Processable* - data is published in a structured manner, to allow automated processing.
6. *Non-Discriminatory* - data is available for all to use, without requiring any registration.

¹³ See: <https://opengovdata.org/>.

7. *Non-Proprietary* - data is published in a format which is not controlled exclusively by a single entity.

8. *Licence-Free* - other than allowing for reasonable privacy, security and privilege restrictions, data is not subject to any limitations on its use due to copyright, patent, trademark or trade secret regulations.

The Sunlight Foundation (2017) subsequently updated the Sebastopol list in 2010 to permit commonly held formats in addition to open formats and added the following two items.

9. *Permanence* - available in perpetuity.

10. *Usage costs* - should be free.

Berners-Lee (2009) set out an alternative approach using the simple 5-Star scheme to gauge and advertise the openness of open linked data. The rating is more tightly focused on the technical aspects of open datasets with respect to format, metadata and linked data:

1-Star = available on the web (whatever format) but with an open licence, e.g., PDF

2-Star = available as machine-readable, structured data, e.g., XLS

3-Star = available in non-proprietary machine-readable, structured format, e.g., CSV

4-Star = above plus use W3C open standards W3C (RDF and SPARQL) to identify things

5-Star = above plus link data to other people's data to provide context

The 5-Star scheme presents a number of fundamental weaknesses in defining and measuring the openness of government data. It is silent with respect to granularity and timeliness whilst including advanced attributes that are beyond the competence of ordinary citizens. As a result, for example, one organogram achieved the highest 5 Star rating, but was last updated seven

years ago;¹⁴ an education dataset, also with the highest 5 Star rating, contains aggregated data.¹⁵ Furthermore, data.gov.uk's 'Openness (Five Stars)' report indicated that 44% of datasets were not open at all, scoring '0 Star' because they were HTML; 4% scored '1 Star' (typically PDF); 4% scored '2 Star' (typically XLS); 37% scored '3 Star' (e.g. CSV, XML); no '4-Star'; 0.3% '5-Star' (RDF); and 10% unrated.¹⁶ However, when sampled, confidence in the statistics rapidly vaporised as the categorisation proved very unreliable and inaccurate. The system appeared confounded by resources presented in alternative formats.

An important difference between the Sebastopol principles and the 5-Star scheme is that the latter is more explicit in recognising degrees of openness. The ODI similarly acknowledges that open data programmes can be a journey and conceptualises levels of openness in two ways. These are: i) the ODI's spectrum of openness is based on lawful accessibility with closed data at one end and fully open data at the other;¹⁷ and ii) the ODI has a four level certification (badge) system similar in style to the 5-Star scheme.¹⁸ The ODI lists 152,910 certified data resources produced by government and non-government entities in 17 countries. 11,620 of these are from the UK: 11,516 are bronze, 98 silver, 4 gold and 2 are platinum. Like the 5-Star scheme, the ODI system is silent with respect to granularity and timeliness. For example, the only two platinum rated datasets are both one-off publications, they are both historic (2011 and 1974) and one contains aggregated data.¹⁹ A unique aspect of the ODI scheme is its emphasis on the

¹⁴ See: <https://data.gov.uk/dataset/e53b480c-e959-4a83-aba5-f146663905d2/organogram-for-government-offices>.

¹⁵ See: <https://data.gov.uk/dataset/72e2ec19-31f3-4663-a1c7-b41eeae01ea1/elective-home-education-camden>.

¹⁶ <https://data.gov.uk/data/report>.

¹⁷ See: <https://theodi.org/data-spectrum>.

¹⁸ See: <https://certificates.theodi.org/en/about/badgelevels>.

¹⁹ See: <https://certificates.theodi.org/en/datasets/1621/certificate> - *European Quality of Life Time Series, 2007 and 2011*; <https://certificates.theodi.org/en/datasets/1709/certificate> - *Family Life and Work Experience Before 1918, 1870-1973*.

capacity of the publisher to support its interaction with users. Consequently, each badge level signals the level of an *organisation's* attainment in publishing a particular dataset. The approach reflects Gurstein's (2013) plea that OGD should be regarded as a service rather than purely as a product.

Bronze = data is openly licensed, no restrictions, accessible and legally reusable.

Silver = as above plus: data is documented in a machine readable format, reliable and offers ongoing support from the publisher via a dedicated communication channel.

Gold = as above plus: in an open machine readable format, has guaranteed regular updates, offers greater support, documentation, and includes a machine readable rights statement.

Platinum = as above plus: has machine readable provenance documentation, uses unique identifiers, publisher has a communications team offering support.

2.5 Measuring openness – current assessment frameworks

Recognising the conceptual distinction between defining and measuring openness, a number of scholars and activists have set out a variety of assessment frameworks.²⁰ Metrics that simply count the number of resources released, are inadequate indicators of the openness of governments and their OGD programmes (Dawes et al., 2016; Harrison et al., 2012). However, there is presently no consensus as to which additional dimensions should be included (Arcelus, 2012; Barry and Bannister, 2014; Bogdanović-Dinić et al., 2014). Attard et al. (2015) provide a systematic review of assessment frameworks. The approaches differ, partly as a consequence of the aspects of OGD selected by the investigators, and partly driven by the selected unit of analysis (for example, purposes or technical features; national governments, portals, websites or datasets). The majority of these techniques incorporate either the 5-Star scheme or

²⁰ See: Attard et al. (2015) for a broad, systematic review of the literature.

Sebastopol principles (Attard et al., 2015). Veljković et al. (2014), for instance, set out a complex, weighted method for assessing the openness and maturity trajectory of national governments. Their OpenGovB model is a combination of indices covering third party corruption, data transparency, user interaction, the presence of key data categories, and the openness of sample datasets according to the Sebastopol principles. Lourenço (2015) manually browsed seven national OGD portals, including data.gov.uk, to qualitatively assess the structural capacity of the portals to deliver the transparency and accountability objectives. The project did not seek to quantify the degree of openness of the datasets within the portals. However, the research was conducted from an ordinary citizen perspective using parameters covering quality, completeness, access and visibility, usability and comprehensibility, timeliness, value and usefulness, and granularity. It included assessments of the portals' interactions with users.

Kubler et al. (2018) used the eGovOI model developed by Veljković et al. (2014) to automate the assessment of portals and websites based on the quality of metadata. Whilst their frameworks may have value in assessing advanced open data qualities that one may expect with 4- or 5-Star datasets, they acknowledged that it cannot capture basic datasets without relevant metadata. Thorsby et al. (2017) developed a mixed qualitative and quantitative method for scoring open data portals in 30 American cities using five scales: the Open Data Portal Index (ODPI) assesses the organisation of portals and the interaction with users; the Data Content Index (DCI) examines the presence of different topics; the Overall Index combines the ODPI and DCI; the Number of Datasets and the Number of Datasets per 100,000 in the population. The project did not test the openness of datasets as the method assumed that resources in the portals were open in compliance with the ten Sunlight Foundation principles (Sunlight Foundation, 2017).

Chatfield and Reddick (2017) conducted a longitudinal study of Australian websites by assessing (1) the adoption of open data policy, (2) the number of datasets published, (3) the prevalence of machine-readable data formats, and (4) the provision of user engagement services, such as hackathons and analytical tools. However just one of the parameters (machine-readability) tested the openness of the datasets. Drilling deeper into the contents of datasets, Vetrò et al. (2016), developed a framework for measuring the quality characteristics of open datasets. The dimensions assessed included, for example, openness on the 5-Star scale, the presence of metadata and the accuracy of cell contents.

It appears that none of these research measurement frameworks have been widely adopted by practitioners, if at all. In practice, despite their limitations, the 5-star and ODI certification systems are the most prevalent schemes that both define and measure the openness of datasets. The two enduring national government level measures are the Open Data Barometer (ODB) and the Global Open Data Index (GODI). The ODB, developed by the World Wide Web Foundation measures the “...true prevalence and impact of open data initiatives around the world.”²¹ The methodology is based on the ten Sunlight Foundation principles (Sunlight Foundation, 2017). The results are presented as a scorecard in terms of readiness, implementation and impact. Implementation is actually a measure of data openness. It assesses the compliance of a sample datasets drawn from 15 key data categories with the following 10 legal and technical dimensions.

1. Does the data exist?
2. Is it available online from government in any form?
3. Is the dataset provided in machine-readable and reusable formats?

²¹ See: <https://opendatabarometer.org/barometer/>.

4. Is the machine-readable and reusable data available as a whole?
5. Is the dataset available free of charge?
6. Is the data openly licensed?
7. Is the dataset up to date?
8. Is the dataset being kept regularly updated?
9. Was it easy to find information about this dataset?
10. Are data identifiers provided for key elements in the dataset?

The GODI produced by Open Knowledge International is very similar to the ODB, providing a national level benchmark of the openness of government data.²² The results are published on a six dimension scorecard, also measuring the openness of a small sample of key datasets in 15 data categories:

1. It's openly licensed
2. It's up to date
3. It's in an open and machine-readable format
4. It's publicly available
5. It's downloadable at once
6. It's available free of charge

The ODB ranks the UK at number one on the country list. It is the only country enjoying the maximum data implementation (or dataset openness) score of 100. The GODI ranks the UK in second place with an overall score of 79%. The most obvious discrepancy between the two UK results is in relation to the election data. The ODB scores the UK's election data at 95 (out of 100); the GODI records the election data at 0%, that is, a complete failure. The two indices may be adequate in providing an overall impression of the performance of national governments, but this divergence does raise methodological and reliability questions. Neither of these two

²² See: <https://index.okfn.org/>.

indicators answer the present research question as they are based on a very small sample of key data resources. They do not measure the depth of data openness by estimating the number of compliant datasets in each territory.

3. Research Method

Following our review, we discovered that existing prevalent measures of openness of data are inadequate for our evaluation. The commonly used 5-Star and ODI schemes are flawed because they omit important dimensions, especially granularity and timeliness. The ODI scheme is also burdened by the entanglement of data and publisher characteristics. The GODI and ODB frameworks are more suitable as they are based on the Sebastopol principles, but they too are silent with respect to granularity. In short, the current openness assessment frameworks assess various aspects of the openness of OGD, such as service and portal qualities, not the openness of datasets from the perspective of the ordinary citizen.

3.1 Constructing the ordinary citizen test

Our approach to the overarching research question - what is the extent of openness of open government data released by the UK government? – was guided by the key principles that open government is enabled by granular data, which is “Usable by All” (G8 Charter, 2013) without advanced technical skills or access to advanced software (Thorsby et al., 2017). Consequently, following Lourenço (2015, p. 324), we set out to assess the openness of individual open government datasets from the perspective of an ‘ordinary citizen’. The ‘ordinary citizen’ is an idealised person who is seeking insight from granular data, has some competence with computers, such as using the Internet and spreadsheets, but does not possess advanced technical skills, such as programming in R or Python, and has no interest in RDF. Our guiding open governance interpretation of OGD focuses on the product and borrows from the DIKW

hierarchy (Rowley, 2007): OGD is the timely, granular data released by governments without any restrictions, it can be processed by any ordinary citizen into information, which contributes to collective knowledge, so that a government may govern more wisely.

Our approach focused on the product (i.e. the dataset) and was unconcerned with the quality of the service delivering the product. The test captured nine essential dimensions that enable datasets to be exploited by a typical, average user, including free access, open format, structured and granular. It excludes the advanced software, data engineering and linking attributes associated with 4-Star and 5-Star ratings. Table 1 sets out how the nine dimensions were interpreted and operationalised for the research. Other descriptive variables were also captured including publisher, title and description, 5-Star rating, ODI certificate and duplicate entries.

Table 1

Ordinary citizen test dimensions

<u>Sebastopol principles</u>	<u>Research dimensions</u>	<u>Recorded variable</u>
Complete	n/a	
Primary	Granular	Granularity
Timely	Recency < 30 months	Latest issue date
Accessible	Published	Published or not
	Operable	Link and dataset functioned
Machine processable	Machine processable	Format(s)
Non-discriminatory	No permission or form filling	Permissions or forms
Non-proprietary	Non-proprietary	Format(s)
Licence-free	Unrestricted licence	Licence type
n/a	Cost-free	Cost

The Sebastopol principle of *completeness* was omitted because it is not a testable attribute of individual datasets. Completeness is a sensible objective, but an ordinary citizen cannot judge whether a government has published all its data, as in the Sebastopol principles; or even, as in the Sunlight Foundation (2017) version, all its data related to a particular topic. Furthermore, the lack of completeness is a product of other dimensions, for example, the absence of granular

data in aggregated statistics, the absence of current data due to long delays in publication, unpublished data, and inoperable resources. Thus completeness of the UK government's OGD efforts is at the core of the present research enquiry, and the results of testing a sample of datasets serves to shed a light on that completeness.

Primary was interpreted as un-aggregated, indivisible granular data. All the datasets in both the most popular and random samples which could be accessed and viewed, or failing that, had adequate descriptions, were carefully categorised according to our 'granularity scale' set out below. The scale is intended to broadly correlate with the value of the data to civic hackers: the higher the granularity, the greater the user-value.

- Report – Highly summarised data at the lowest resolution included in or supporting written reports, tables may also be published separately in spreadsheet format.
- Summary – Summarised data broken down into a small number of categories, independent of reports, but may have written supporting explanations.
- Category – Data broken down into medium to high number of categories, but above the granular level.
- Granular – Data at the highest resolution and cannot be sensibly divided further, e.g., person, vehicle, location of building, and has not been analysed by intermediaries.

Timely is important because data which reflects current social conditions and government practices has more open government utility than historic data (Vetrò et al., 2016). It is, however, subjectively defined in the literature, using qualifiers such as 'as soon as possible' (Sebastopol principles), continuous release (Lourenço, 2015) and the presence of updates (Vetrò et al., 2016). The Sunlight Foundation (2017) acknowledges that timeliness is contingent on the

nature of the data, and encourages real-time release to maximise its utility, a robust criterion that would disqualify virtually all OGD publications. We operationalised *Timely* as ‘recency’, that is, datasets published within the previous 30 months qualified as open. This is somewhat generous, but it ensured that publications covering the most recent annual reporting period were captured.

Accessible became ‘published’ and ‘operable’. An initial foray into the data.gov.uk portal revealed that some advertised datasets are not published at all, whilst others are inoperable due to broken links, persistent failure to download and fatal opening errors. *Machine processable* data is presented in a structured format that can be read and processed by computers. *Non-discriminatory* was interpreted as datasets that do not require access permissions or completing registration forms. *Non-proprietary* data is published in open software formats, for example, CSV and GeoJSON. Where data was presented in alternative formats, the most open was used to categorise the dataset, for example CSV rather than XLS and XLS rather than PDF. For *licence-free*, an unrestricted licence qualified as open: the vast majority of resources on data.gov.uk declare some form of licence, mainly the Open Government Licence (OGL). Finally, OGD should be free of any charges. Licence-free does not necessarily mean cost-free as it is possible that data subject to licence restrictions may be cost free, whilst gateway fees may be charged to access data with unrestricted licence conditions. The Sebastopol group later acknowledged that datasets should be free of any charges²³ and it has become an accepted principle (Attard et al., 2015; Bates, 2014; HM Government, 2016; Sunlight Foundation, 2017).

3.2 Sampling data.gov.uk portal

²³ See: <https://opengovdata.org/>.

The data.gov.uk website provides statistics on the number of visits, views and downloads for each of its datasets since July 2012.²⁴ Our analysis is based on the statistics on the 22nd of May 2017. The file includes 36,263 datasets deposited up to this date. The portal warns that the statistics exclude datasets for which the number of views/visits is ‘tiny’ and only count users with JavaScript enabled, which excludes access by web crawlers and Application Programming Interface (API) calls.²⁵ It also excludes access via departmental websites. Despite these limitations, these usage statistics are indications of the popularity of datasets, particularly for ordinary citizens who are unfamiliar with web crawlers and APIs. Data.gov.uk also produces some reports for publishers, which provide various statistics including the extent of 5-Star openness, broken datasets and unreleased datasets. These were reviewed and compared to our findings.²⁶

Two samples were taken and manually examined by a single researcher to assess the openness of datasets on the data.gov.uk website. The second researcher then reviewed and verified the results. A purposive sample of the most popular datasets was examined to gauge the level of openness most likely to be experienced by ordinary citizens through their encounters with the data catalogue. The usage statistics derived from the data.gov.uk usage file indicate a heavy skew with a small number of resources attracting the most attention (Table 2 and Table 3). The ‘popular sample’ included all the most frequently accessed datasets with more than 10,000 views or more than 3,000 downloads. The criteria produced 136 datasets (0.38% of the population), which represented 49% of all downloads and 39% of all views.

²⁴ See: <https://data.gov.uk/data/site-usage#totals>.

²⁵ See: <https://data.gov.uk/data/site-usage/dataset>.

²⁶ See: <https://data.gov.uk/data/report>.

Table 2

Popularity of datasets – number of views.

# views	# datasets	% datasets	S views	% views	cum% views
>60k	14	0.039%	1,243,655	14.1%	14.1%
30-60k	16	0.044%	680,453	7.7%	21.8%
20-30k	16	0.044%	388,035	4.4%	26.1%
10-20k	63	0.174%	877,070	9.9%	36.1%
5-10k	123	0.339%	859,139	9.7%	45.8%
3-5k	173	0.477%	653,813	7.4%	53.2%
2-3k	209	0.576%	505,954	5.7%	58.9%
1-2k	617	1.701%	861,264	9.7%	68.6%
500-999	1,136	3.133%	789,969	8.9%	77.6%
250-499	1,907	5.259%	670,124	7.6%	85.2%
100-249	4,281	11.805%	673,366	7.6%	92.8%
1-99	27,708	76.408%	639,503	7.2%	100.0%
0	0	0.000%	0	0.0%	100.0%
Total	36,263	100%	8,842,345	100%	

0.48% of resources attract 53% of views

Table 3

Popularity of datasets – number of downloads.

# d/loads	# datasets	% datasets	Sd/loads	% d/loads	cum% d/loads
>60k	1	0.003%	95,674	4.1%	4.1%
30-60k	4	0.011%	102,630	4.4%	8.5%
20-30k	7	0.019%	236,937	10.1%	18.6%
10-20k	16	0.044%	208,545	8.9%	27.6%
5-10k	24	0.066%	160,939	6.9%	34.5%
3-5k	51	0.141%	189,956	8.1%	42.6%
2-3k	58	0.160%	129,634	5.6%	48.1%
1-2k	164	0.452%	219,431	9.4%	57.5%
500-999	326	0.899%	218,474	9.4%	66.9%
250-499	493	1.360%	175,090	7.5%	74.4%
100-249	1,290	3.557%	199,662	8.6%	83.0%
1-99	21,975	60.599%	380,485	16.3%	99.2%
0	11,854	32.689%	17,589	0.8%	100.0%
Total	36,263	100%	2,335,046	100%	

0.14% of datasets contribute 43% of downloads

The second sample was a random selection of 264 datasets used to examine the openness of datasets irrespective of their popularity (0.75% of the population, 4% of views and 4% of downloads). Thus, the total sample frame was 400 datasets. We also found 7 duplicate entries on data.gov.uk which have been excluded from the present analysis. The margin of error on the

(n=264) random sample is +/-6% at the 95% confidence level. The datasets were manually approached, simulating the manner of an ordinary citizen. The webpage for each dataset was located and the relevant descriptive metadata and attributes published on these webpages were recorded. The data files were then accessed and viewed whenever possible and observations recorded. This step excluded the inaccessible resources that were unpublished, some files failed to open, others linked to websites which required access permissions or charged fees. All these events were noted.

4. Findings

4.1 Openness of datasets

For the most popular sample of datasets (n=136), just 20 (15%) are open from the perspective of the ordinary citizen, equating to 0.05% of the population of datasets on the data.gov.uk catalogue. The 20 open datasets are dominated by organisation charts (6), transport data (6) and map data (4). This represents the level of openness of UK government data that is most frequently experienced by users. Similarly, for the random sample of datasets (n=264), 34 (13%) are open based on the ordinary citizen criteria. Table 4 identifies the range of failure causes or barriers to the lack of openness of these datasets in frequency order. The total number of failures exceeds the sample sizes because many datasets suffer from multiple failure types. We counted a total of 247 barriers across the 116 non-open datasets in the most popular sample, and 482 barriers across the 230 non-open random sample datasets. The average in both samples is 2 per dataset.

The most common barrier in both samples is aggregation: approximately 6 out of 10 datasets contain aggregated information, not granular data. Further granularity findings are presented later in this section. The second most common barrier in both samples is aged data: 4 out of 10

datasets are more than 30 months old. Considering that timeliness is a key dimension of OGD (Sunlight Foundation, 2017), the effect of timeliness on openness was tested by adjusting the recency criterion in 6 months steps from 6 months to 42 months. Figure 1 shows that 4% of datasets would be regarded as open if the definition of OGD demanded that data should be less than 6 months old. This rises to 13% if the recency criterion is set to capture all datasets that are less than 30 months old, and to 16% if timeliness is entirely ignored (recency > 42 months).

The overall level of inaccessibility in both samples is high. A total of 27 (19%) popular datasets and 81 (31%) random sample datasets were not readily accessible to ordinary members of the public as they were unpublished (5% / 17%), inoperable (10% / 10%), required permissions or form filling (3% / 3%), or charged a fee (1%, 0.4%).

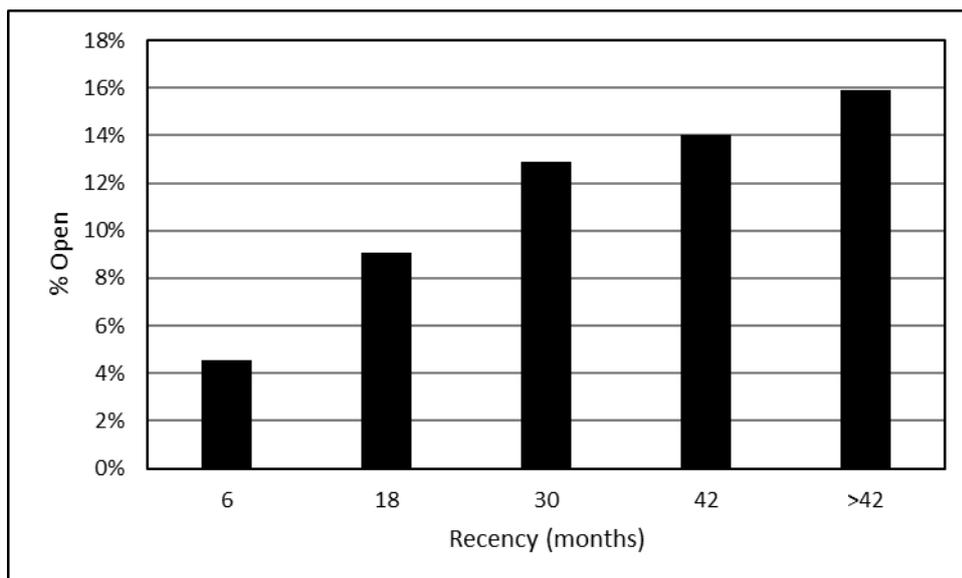
Table 4

Barriers to openness of UK open government data.

Research dimension	Failure cause	Most popular sample		Random sample	
		Frequency	%	Frequency	%
Granular	Aggregation, not granular	79	58%	175	66%
Recency < 30 months	Aged - over 30 months old	62	45%	103	39%
Machine processable	Not structured	36	26%	27	10%
Non-proprietary	Not open format	29	21%	48	18%
Unrestricted licence	Licence restrictions	14	10%	48	18%
Operable	Inoperable, failed to function	14	10%	26	10%
Published	Unpublished / not available	7	5%	46	17%
No permission or form filling	Require permission or forms	4	3%	8	3%
Cost-free	Require fee	2	1%	1	0.4%
Total		247		482	

Figure 1

Timeliness – the effect of the recency criterion



Site statistics from data.gov.uk’s publisher reports were assessed.²⁷ However, the portal’s own measures are not directly comparable with our results. For example, an ‘Openness (Five Stars)’ report based on the 5-Star rating indicated that 44% of datasets scored ‘0 Star’, mainly because they were HTML, and 4% scored ‘1 Star’, mainly because they were PDF.²⁸ By comparison, the ‘not structured’ category in our sample included 6% in PDF and 4% in HTML web pages. Whereas the data.gov.uk categorisation system appeared confounded by multiple formats for individual resources, our approach always selected the most open format. Since the portal’s refresh in March 2018, it no longer advertises 5-Star ratings (data.gov.uk, u.d.). Unfortunately, the site’s licence report was unusable: it indicated over 2,000 different licences in use, but this variation is driven by publishers’ comments in a custom term field. The site measures the number of datasets without resources (2%) separately from the number unpublished (7%). We classified all instances where entries were unpublished, not released or otherwise completely absent as unpublished (17%) since the reasons for their absence could not be reliably determined. The website reported 14% ‘datasets with broken links’, whereas our ‘inoperable’ category was concerned with broken links, failed downloads and failure to open (10%).

²⁷ <https://data.gov.uk/data/report>.

²⁸ <https://data.gov.uk/data/report/openness>.

4.2 Granularity scale

Granularity is considered as the most prominent barrier to openness due to its crucial function in enabling the grand purposes of OGD – open government. Yet, it is glossed over in some assessment frameworks, such as those in G8 (2013) and Dawes et al. (2016). Table 5 sets out the granularity results of the two samples. 47 (35%) of the most popular resources are granular. An important factor in the granularity barrier is the privacy of citizens. Just three of the 47 granular datasets directly relate to members of the public. The Land Registry Price Paid Data²⁹ and the Road Traffic Accidents³⁰ datasets are anonymous, though individuals can be re-identified by reference to other sources. The Bona Vacantia Unclaimed Estates³¹ identifies deceased persons. Similarly, 96 (36%) of the random sample datasets are at the granular resolution. Again reflecting privacy concerns, just one small, published dataset within this group directly related to members of the public: the Court Martial Results from the Military Court Centres, which lists 2,759 sentences imposed on military personnel from 2010 to 2015.³² The other 95 datasets are overwhelmingly dominated by map and location data.

Table 5

Granularity of UK open government data.

Granularity	Most popular		Random	
Report	19	14%	52	20%
Summary	11	8%	18	7%
Category	34	25%	59	22%
Granular	47	35%	96	36%
Not assessable	25	18%	39	15%
Total	136		264	

²⁹ See: <https://data.gov.uk/dataset/land-registry-monthly-price-paid-data>.

³⁰ See: <https://data.gov.uk/dataset/road-accidents-safety-data>.

³¹ See: <https://data.gov.uk/dataset/bona-vacantia-estates-advertisements>.

³² See: https://data.gov.uk/dataset/court_martial_results_military_courts.

5. Discussion

5.1 Definition and assessment frameworks

This research did not set out to provide a critique of existing OGD definition and assessment frameworks. However, our findings draw critical attention to how the openness of OGD is defined and measured, particularly with respect to the most prevalent measures, including the 5-Star scheme, the ODI certificates, and the GODI and ODB transnational frameworks.

A key dimension frequently absent, or not explicit mentioned, in these definition and assessment frameworks, is granularity. This lack of explicitness contributes to the conceptual ambiguities around OGD identified by Yu and Robinson (2012). Granularity is also one of the least visible barriers in Table 4, in the sense that it is, in general, not included in descriptions and metadata. Thus, it does not become visible until a resource is downloaded and opened. In some frameworks, granularity is implied, typically in the requirement for machine readability, overlooks the fact that mean values can be just as machine readable as raw data. Both of the most prevalent definitional schemes, the 5-Star scheme and the ODI certificates, suffer from this problem. Indeed, our findings show that the government's use of the 5-Star rating system to report openness is flawed as it is solely based on categorising format types, fails to discriminate between granular and aggregate data, and is unreliable in its operation. It is notable that, following an update in 2018, data.gov.uk no longer advertises 5-Star scheme because it confused the users (data.gov.uk, u.d.), and the ODI certificates have also disappeared without comment. The 5-Star scheme and ODI certificates are also inadequate with respect to timeliness. Timeliness can have a number of subsidiary components, such as the time between data collection and release, and frequency of updates. Although aged data has some value in, for example, trend analyses, it is usually most valuable when it reflects current political, social and economic conditions (Vetrò et al., 2016).

The GODI and ODB frameworks may be adequate in providing an impression of relative performance at a national level. However, they do not measure the extent of openness and include datasets that contain aggregated information, which means they are immune to interrogation and fundamental challenge, for example, GDP and crime statistics. Consequently, a government can create images of openness by focusing its efforts on a few important publications. Our data suggests that the reputation they accord the UK government contradicts reality, exaggerate the openness of its data and hence the openness of its governance. They support the UK government in proclaiming its ‘number one slot’ (Heywood, 2016) as ‘a world leader in making public sector available for re-use’ (UKAuthority, 2018), whilst deflecting attention from the fact that the bulk of its publications are low utility informational resources, not re-usable, granular open data.

Barry and Bannister (2014) complained that the term open data is used loosely and is often poorly defined. It allows governments to negotiate the space between closed and open data for their own purposes (Cole, 2012). Based on our experience, until a firm consensus is reached regarding definitions and assessment frameworks, researchers need to be explicit in the definitions and interpretations they use to guide their individual research programmes. Having evaluated a range of alternative approaches, we suggest that the Sebastopol principles provide a sound analytical reference point for assessing datasets. Based on these principles, researchers could subsequently set out and explain any variations or interpretations associated with their own projects. Our descriptive DIKW definition in section 3.1 and the technical framework set out in Table 1 are firmly based on the Sebastopol principles.

5.2 Implications for OGD and open government

The UK government's actions and its position on the GODI and ODB leader-boards suggest that it is in the vanguard of the OGD movement. Over the past eight years, the politicians have stimulated the initiative by asking for and getting the publication of tens of thousands of resources labelled as open data. Our research found that just over 1 in 10 (13%) government datasets are open from the perspective of the ordinary citizen. The implication is that there are not 43,000 open government datasets as advertised by data.gov.uk. From the perspective of an ordinary citizen, it is closer to 5,500.

Our findings suggest that the main reasons that open resources are not open data is that the content of nearly two-thirds (65%) of the resources is aggregated information as opposed to granular data, 38% are aged data and 30% are not readily accessible. The distinction between data and information as set in the knowledge hierarchy is critical (Rowley, 2007; Tuomi, 1999). Government mediated open information cannot fully engage the analytical capacities of citizens in order to break the interpretive monopoly of the government (Baack, 2015). The quantity of aged data further undermines these principles as it has less value and utility in describing current social conditions and challenging government practices (Sunlight Foundation, 2017; Vetrò et al., 2016). The level of inaccessibility contravenes the core transparency principle behind open government and open government data (G8, 2013; Heald, 2006; McGee and Edwards, 2016). In combination these problems severely neutralise the public's civic hacking capacity (Magalhaes, Roseira and Strover, 2013; Tauberer, 2014). Most of the datasets listed on data.gov.uk can be accessed directly from the platform, but many are in effect catalogue entries that link to other websites. Some of these external websites then raise further hurdles by requiring users to conduct additional searches and then seek access permissions, for example the post code tool, Code-Point Open.³³ Although this research is constructed from the

³³ See: <https://data.gov.uk/dataset/code-point-open1>.

perspective of the ordinary citizen, the findings substantively apply to businesses and researchers because the dominant barriers are lack of granularity and timeliness. These barriers, however, are less problematic for businesses and researchers as they can access alternative, privileged routes, such as primary research, sharing with peers, government datalabs (e.g., HMRC³⁴ and ONS³⁵) and the UK Data Service.³⁶

The barriers to OGD, and thence to more open government, alluded to in this paper have been extensively examined in the literature, see for example Attard et al. (2015) and Barry and Bannister (2014). These barriers range from a defensive culture (Grimmelikhuijsen, 2013) to poor awareness (Zhao and Fan, 2018), lack of management support (Ruijter et al., 2017), lack of measured benefits (Davies, Perini and Alonso, 2013) and privacy risks (Janssen and van den Hoven, 2015). Our findings suggest that privacy concerns have severely limited the number of granular datasets that relate to members of the public. A measure of the government's privacy control framework is that the only one granular dataset, which survived the openness test in the random sample, is a short-anonymised list of sentences imposed on 2,759 military personnel.

We suggest that one consequence of these barriers is that the vast majority of the UK government's open data products attract little public attention: 53% of the views recorded by data.gov.uk are associated with just 0.48% of the publications. The government's initiative appears to be, as Janssen et al. (2012) observe, supply orientated rather than demand driven, and lacks attention to the users' perspective. The result is that a minority of datasets with real potential are drowned in a swamp of irrelevant Zombie data (Gurin, 2014; dos Santos Brito et

³⁴ See: <https://www.gov.uk/government/organisations/hm-revenue-customs/about/research#the-hmrc-datalab>.

³⁵ <https://www.ons.gov.uk/aboutus/whatwedo/paidservices/virtualmicrodatalaboratoryvml>.

³⁶ See: <https://www.ukdataservice.ac.uk/>.

al., 2014). Grimmelikhuijsen (2013) suggested that data about non-government organisations, traffic and transport data adds little value to open government. The materials sampled for this research were similarly dominated by low value administrative (e.g., organograms), transport and location data. Kornberger et al. (2017) described programmes, which produce these kinds of data as displacement activities, designed to avoid open governance; whilst giving the impression of openness. There appears to be few resources in the high utility nexus between high value topics and timely, granular open data, which would encourage innovation or novel insights into government practices. The contribution of the UK government's OGD programme to open government is far more limited than the headline publication statistics suggests. These observations imply that, despite its respected position, the UK government's energies lean more towards impression management than effective open government. It can be confident in doing so by hiding behind the conceptual ambiguities of the open government movement (Yu and Robinson, 2012) and inadequate measures of data openness.

The paucity of high utility resources may go some way to explain why civic participation has remained stubbornly low (Attard et al., 2015) and activists have failed to measure positive outcomes (Davies, Perini and Alonso, 2013; European Commission, 2015; Janssen et al., 2012). This lack of measurable benefits means that OGD remains an experimental endeavour. The implication of our findings is that continuing with current practices may not yield the hoped-for benefits. The big question is whether, given the legal restrictions and other barriers, it can ever live up to its promises. Does the experiment need to be unshackled by a significant, yet unlikely shift in the privacy landscape to enable the release of granular data related to people? It would certainly help if publishers clearly segmented and labelled open resources to distinguish open data from open information.

6. Conclusions

Researchers have previously highlighted the lack of utility of OGD resources in advancing the open government purpose. This paper has quantified a deficiency that significantly contributes to the utility problem in the UK. The effort required limited the random sample size, nevertheless with a margin of error of +/-6% at the 95% confidence level, the results reflect the extent of openness of OGD in the UK with reasonable accuracy. Only a small minority (approximately 1 in 10) of UK government resources portrayed as OGD are actually open from the perspective of the ordinary citizen. The UK government's bold OGD initiative has resulted in a swamp of non-granular, unstructured, aged and frequently inaccessible information that drowns open data. The adage, less is more, would seem appropriate, but one can imagine the political furore triggered by accusations of secrecy should the UK government seek to drain the swamp. The situation reflects a convenient, supply-driven coping strategy which produces few high utility datasets that facilitate the movement's grand purposes. We suggest that an alternative approach would be to clearly signpost the distinction between data and information so that ordinary citizens could more easily find or request relevant high utility, granular datasets. Returning to the Sebastopol principles for a common measure of data openness would bring greater transparency to government claims of transparency.

There will always be moral and legal barriers to releasing high utility datasets due to privacy and security considerations (Janssen and van den Hoven, 2015). The UK government appears to be highly effective in minimising any privacy risks associated with OGS. However, a high quantity of informational publications that creates the illusion of open government does not compensate for these restrictions. Despite its leading OGD position, the UK government appears to be locked into an impression management strategy that exploits the conceptual ambiguities around open government and relies on a volume supply of low utility products. Our

findings provide a measure of these ambiguities, particularly with respect to granularity and timeliness.

User engagement is clearly problematic with the great majority of publications attracting hardly any attention. It is also notable that data.gov.uk has switched off its commenting and forum features due to inactivity (data.gov.uk, u.d.). The government could borrow practices from the business sector to get to know its customers, to understand their product demands and what they value. After all, the premise of the OGD movement is that citizen-users deliver its grand purposes. Further academic research into what topics users value would help steer government practices. At the present level of market maturity, the quality of the product and the quality of the service needs to be viewed separately: the service is less salient if the product is routinely non-compliant. OGD assessment frameworks too need to separate the issues. Using the Sebastopol principles as the basis for an ordinary citizen test is an effective way of assessing the openness quality of datasets.

It is important to re-emphasise that this research has not evaluated the performance of the data.gov.uk portal. The research is also based on a snapshot of the data resources listed on the data.gov.uk portal in May 2017. The contents of the portal change continuously. A brief examination of some of the research datasets following the site's refresh in early 2018 showed that some of the metadata had disappeared, making the exercise more difficult to replicate. Timeliness is a subjective criterion. We set the test limit to datasets last updated within the previous 30 months, which does not capture the period between data collection and last publication. Nevertheless, we have shown that the extent of openness rapidly declines as one approaches real-time. Other than privacy restrictions, this paper has not examined the causes behind the lack of granularity, timeliness and non-discrimination. A forthcoming interview-

based paper will explore these issues. Further research into how much civic hacking feedback the government receives, and whether it listens, would provide insight into whether one of the grand purposes of OGD – open government – is meaningful. This research was conducted in the context of the prevalent technologies and legal requirements. It would be useful to assess whether the General Data Protection Regulation (GDPR) has any impact on the behaviour of government agencies regarding the release of OGD. Further, comparative studies could also be carried out in other countries to assess the extent of openness of their datasets. Considering that the UK is regarded as a leader in the OGD movement, the deficiencies highlighted here may be amplified in other territories.

Acknowledgement

This research was supported by the EPSRC project Data Release – Trust, Identity, Privacy and Security (EP/N27825/1).

References

- Altman, M, Wood, A., O'Brien, D., Vadhan, S., & Gasser, U. (2015). Towards a modern approach to privacy-aware government data releases. *Berkeley Technology Law Journal*, 30(3), 1967-2072.
- Arcelus, J. (2012). Framework for useful transparency websites for citizens. In *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance* (pp. 83-86). ACM.
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399-418.
- Axelsson, A., & Schroeder, R. (2009). Making it open and keeping it safe: e-enabled data-sharing in Sweden. *Acta Sociologica*, 52(3), 213-226.

- Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, 2(2). DOI: org/10.1177/2053951715594634
- Ballingall, R. (2011). Is there a case for limiting democratic openness? *Public Policy and Governance Review*, 2(2): 52-62.
- Barry, E., & Bannister, F. (2014). Barriers to open data release: A view from the top. *Information Polity*, 19(1, 2), 129-152.
- Bates, J. (2014). The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom. *Government Information Quarterly*, 31(3), 388-395.
- Berners-Lee, T. (2009). *Linked data*. Available at: <http://www.w3.org/DesignIssues/LinkedData.html> (accessed Nov. 2016).
- Bogdanović-Dinić, S., Veljković, N., & Stoimenov, L. (2014). How open are public government data? An assessment of seven open data portals. In M.P. Rodríguez-Bolívar (Ed.), *Measuring E-government Efficiency, Public Administration and Information Technology*, vol. 5. (pp. 25–44). New York: Springer.
- Chan, C. M. (2013). From open data to open innovation strategies: Creating e-services using open government data. In *2013 46th Hawaii International Conference on System Sciences* (pp. 1890-1899). IEEE.
- Chatfield, A. T., & Reddick, C. G. (2017). A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments. *Government Information Quarterly*, 34(2), 231-243.
- Cole, R. (2012). Some observations on the practice of “open data” as opposed to its promise. *The Journal of Community Informatics*, 8(2).
- Data.gov.uk (u.d.). *Update about changes to data.gov.uk*. Available at: <https://data.gov.uk/site-changes> (accessed Oct. 2018).
- Davies, T., & Perini, F. (2016). Researching the emerging impacts of open data: revisiting the ODCC conceptual framework. *The Journal of Community Informatics*, 12(2): 148-178.
- Davies, T., Perini, F., & Alonso, J. (2013). *Researching the emerging impacts of open data*. Open Data Research Network. Available at: <http://www.opendataresearch.org/reports> (accessed June 2017).
- Dawes, S. S., Vidasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1), 15-27.
- dos Santos Brito, K., da Silva Costa, M. A., Garcia, V. C., & de Lemos Meira, S. R. (2014). Brazilian government open data: implementation, challenges, and potential opportunities. In *Proceedings of the 15th Annual International Conference on Digital Government Research* (pp. 11-16).

- Eaves, D. (2009). The three laws of open government data. Available at: <https://eaves.ca/2009/09/30/three-law-of-open-government-data/>
- European Commission (2015). *Creating Value through Open Data*. Available at: https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf (accessed Feb. 2017).
- G8 (2013). *Open Data Charter*. Available at: <https://www.gov.uk/government/publications/open-data-charter> (accessed Feb. 2017).
- Gigler S, Custer S., & Rahemtulla H. (2011). *Realizing the vision of open government data*. Washington, DC: Open Development Technology Alliance. Available at: <http://www.scribd.com/doc/75642397/Realizing-the-Vision-of-Open-Government-Data-Long-Version-Opportunities-Challenges-and-Pitfalls>.
- Grimmelikhuijsen, S. (2013). A good man but a bad wizard. About the limits and future of transparency of democratic governments. *Information Polity*, 17(3), 293-302.
- Gurin, J. (2014). Open governments, open data: A new lever for transparency, citizen engagement, and economic growth. *SAIS Review of International Affairs*, 34(1), 71-82.
- Gurstein, M. (2013). 'Should "Open Government Data" be a product or a service (and why does it matter?)', Gurstein's Community Informatics, 3 February. Available at: <http://gurstein.wordpress.com/2013/02/03/is-open-government-data-a-product-or-a-service-and-why-does-it-matter/> (accessed April 2017).
- Harrison, T.M., Guerrero, S., Burke, G.B., Cook, M., Cresswell, A., Helbig, N., Hrdinová, J., & Pardo, T. (2012). Open government and e-government: Democratic challenges from a public value perspective. *Information Polity*, 17(2), 83-97.
- Heald, D. (2006). Varieties of transparency. *Proceedings of the British Academy*, 135, 25-43.
- Heimstädt, M. (2017). Openwashing: A decoupling perspective on organizational transparency. *Technological Forecasting and Social Change*, 125, 77-86.
- Heywood, J. (2016). *Open data – the revolution is here*. Available at: <https://civilservice.blog.gov.uk/2016/04/27/open-data-the-revolution-is-here/> (accessed Oct. 2018).
- HM Government (2016). *UK Open Government National Action Plan 2016-18*. Available at: <https://www.gov.uk/government/publications/uk-open-government-national-action-plan-2016-18/uk-open-government-national-action-plan-2016-18> (accessed March 2017).
- Janssen, K. (2012). Open government data: Right to information 2.0 or its rollback version? ICRI Research Paper (8). Available at: <file:///C:/Users/user/Downloads/SSRN-id2152566.pdf> (accessed Oct. 2017).
- Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, 32, 363-368.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258-268.

- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3): 262-267.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, 31(3), 471-481.
- Kornberger, M., Meyer, R., Brandtner, C., & Hollerer, M. (2017). When bureaucracy meets the crowd: Studying 'Open Government' in the Vienna city administration. *Organization Studies*, 38(2), 179-200.
- Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2017). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35(1), 13-29.
- Kučera, J., Chlapek, D. and Nečaský, M. (2013). Open government data catalogs: Current approaches and quality perspective. In *International conference on electronic government and the information systems perspective* (pp. 152-166). Springer, Berlin, Heidelberg.
- Lagoze, C. (2014). Big Data, data integrity, and the fracturing of the control zone. *Big Data & Society*, 1(2). DOI:org/10.1177/2053951714558281
- Leonelli, S. (2012). Introduction: making sense of data-driven research in the biological and biomedical sciences. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(1): 1-3.
- Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, 32(3), 323-332.
- Magalhaes, G., Roseira C., & Strover, S. (2013). Open government data intermediaries. In: *7th international conference on theory and practice of electronic governance, ICEGOV 2013*, Seoul, South Korea, 22–25 October 2013, pp. 330– 333.
- Mayernik, M. (2017). Open data: Accountability and transparency. *Big Data & Society*, 1-5. DOI: 10.1177/2053951717718853
- McGee, R. (2011). Apparently transparent: Does accountability deliver? *The Broker*. Available at: <http://www.thebrokeronline.eu/Articles/Apparently-transparent>.
- McGee, R., & Edwards, D. (2016). Opening Governance - Change, Continuity and Conceptual Ambiguity. *IDS Bulletin*, 47(1): 1–22.
- McGee, R., & Gaventa, J. (2011). *Shifting power? Assessing the impact of transparency and accountability initiatives*. Institute of Development Studies. Available at: <http://www.ids.ac.uk/publication/shifting-power-assessing-the-impact-of-transparency-and-accountability-initiatives> (accessed April 2017).
- Millard, J. (2015). Open governance systems: Doing more with more. *Government Information Quarterly*, <http://dx.doi.org/10.1016/j.giq.2015.08.003>
- Parycek, P., & Sachs, M. (2010). Open Government – Information Flow in Web 2.0. *European Journal of ePractice*, 9. Available at:

<http://www.etudasportal.gov.hu/download/attachments/7995452/European+Journal+epractice+Volume+9.5.pdf> (accessed March 2017).

Robinson, D., Yu, H., Zeller, W.P., & Felten, E.W. (2009). Government data and the invisible hand. *Yale Journal of Law & Technology*, 11, 160.

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2): 163-180.

Ruijter, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, 34(1), 45-52.

Schrock, A., & Shaffer, G. (2017). Data ideologies of an interested public: A study of grassroots open government data intermediaries. *Big Data & Society*, 4(1). DOI: [org/10.1177/2053951717690750](https://doi.org/10.1177/2053951717690750)

Sunlight Foundation (2017). *Ten Principles for Opening Up Government Information*. Available at: <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/>

Tauberer, J. (2014). *Open government data* (2nd ed.). Available at: <https://opengovdata.io/2014/civic-hacking/> (accessed March 2017).

Thorsby, J., Stowers, G. N., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly*, 34(1), 53-61.

Tuomi, I. (1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems*, 16(3): 103–117.

UKAuthority (2018, Sept 24). *UK ranked joint top of Open Data Barometer*. Available at: <https://www.ukauthority.com/articles/uk-ranked-joint-top-of-open-data-barometer/> (accessed Oct. 2018).

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, 31(2), 278-290.

Verhulst, S., & Young, A. (2016). *Open data impact: when demand and supply meet*. New York: The GovLab. Available at: <http://odimpact.org/key-findings.html> (accessed June 2017).

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337.

W3C (2004). RDF Primer. Available at: <https://www.w3.org/TR/rdf-primer/> (accessed May 2016).

White House (2009). *Open Government Directive*. Executive Office of the President. Available at: <https://obamawhitehouse.archives.gov/open/documents/open-government-directive> (accessed June 2017).

Yiu, C. (2012). *A Right to Data: Fulfilling the Promise of Open Public Data in the UK*. Policy Exchange. Available at: <http://www.policyexchange.org.uk/publications/category/item/a-right-to-data-fulfilling-the-promise-of-open-public-data-in-the-uk> (accessed Oct. 2017).

Yu, H. and Robinson, D. (2012). The New Ambiguity of “Open Government”. *UCLA Law Review Discourse*, 59, 178-208.

Zhao, Y., & Fan, B. (2018). Exploring open government data capacity of government agency: Based on the resource-based theory. *Government Information Quarterly*, 35(1), 1-12.